Shifting Gears:
Gearing Up to Get Into the Flow

Ricky Erway and Jennifer Schaffner

Program officers
OCLC Programs and Research

Shifting Gears: Gearing Up to Get Into the Flow
Ricky Erway and Jennifer Schaffner, for OCLC Programs and Research

Suggested citation:
Erway, Ricky, and Jennifer Schaffner. 2007. Shifting Gears: Gearing Up to Get Into the Flow.
Report produced by OCLC Programs and Research. Published online at:
www.oclc.org/programs/publications/reports/2007-02.pdf

As a community, we have spent more than two decades painstakingly pursuing the highest quality in our digitization of primary resources. Through Google Books, the Open Content Alliance, and similar efforts, book collections are flying off the shelves and finding their way to users in digital form. In a world where it is increasingly felt that if it's not online it doesn't exist, we need to make sure that our users are exposed to the wealth of information in special collections.

Katie Hafner, in the New York Times article, "History, Digitized (and Abridged)," quotes Edward L. Ayers, historian and now president of the University of Richmond, as saying:

> There's an illusion being created that all the world's knowledge is on the Web, but we haven't begun to glimpse what is out there in local archives and libraries. Material that is not digitized risks being neglected as it would not have been in the past, virtually lost to the great majority of potential users.[1]

Scaling up digitization of special collections (here defined as non-book collections, such as photographs, manuscripts, pamphlets, minerals, insects, or maps) will compel us to temper our historical emphasis on quality with the recognition that large quantities of digitized special collections materials will better serve our users. This will require us to revisit our procedures and policies. Should we be digitizing for both preservation and access, or optimizing procedures primarily for access? How can our selection approaches help us maximize both throughput and impact? Have projects produced reusable infrastructures? What is the appropriate level of description for online materials? How can we make smart partnership agreements in order to build a collective collection that will be valued by a broad audience?

This essay endeavors—like the "Digitization Matters" forum that inspired it (see Appendix)—to challenge its audience to take a fresh look at approaches to extending access to the special collections in libraries, archives, and museums. The forum speakers were asked to be provocative, not to represent what they or their institutions have done, but to focus on ideas for significantly increasing the scale of our digitization activities. Because of this somewhat unusual approach—and because so many of the ideas were further developed in the open discussions—we're not providing attribution for each idea, but rather summarizing the outcomes of the forum as a whole.

The essay, like the forum, focuses on digitization and related processes, but intentionally does not encompass technical specifications for various formats, born digital materials, nor rights issues (which warrant similar essays for each topic). It intends to be provocative. Not all of the ideas presented here will apply to a particular situation, but hopefully they will stimulate consideration of appropriate ways to move forward.

Special collections are stuck in an eddy, while the mass of digitized books drift by in the current of the mainstream. We need to jump into the flow or risk being left high and dry.

1.   Access vs. preservation—Access wins!

Many of our digital initiatives have stressed the importance of preservation, leaving access as an afterthought (the idea being if you capture preservation-quality; you can always derive an access copy). In reality, due to the very special nature of these often unique materials, we will always preserve the originals to the best of our ability. In light of recent programs for the mass digitization of books, if special collections and their funding continue to be marginalized, our administrations may not keep us around to attend to the originals.

In the past, we've soothed our doubts by repeating the mantra, "we'll only get one chance to do it, so it's got to be done right." Experience has shown that that is not in fact the case. Often we do go back when the technology improves or when we better understand our users' needs. We need to put on our helmets now and go for the biggest bang for the buck in terms of access.

2.   Selection has already been done

We often speculate what would be appropriate for scholars and what could attract the general public. Scholars can make new discoveries in just about any materials they can access and the general public often eagerly delves into that which we thought might not interest them. We've spent a lot of time guessing what will be useful to our users, but we need to spend more time learning from our users (and listening when they tell us) what they want.

There are three reasonable means to determine what to digitize:

A. Scan as materials are accessioned. If we've carefully appraised and acquired collections that are suited to our missions, those collections are worthy of digitization.

i. For collections known to be of great interest and that lend themselves to digitization, think about scanning the entire collection rather than making decisions about which bits.

ii. Otherwise think about scanning the first chunk, or a box of correspondence, or samples from several parts of the collection.

B. Scan on demand. Scholarly and casual use of our collections identify materials that are likely to be used again. As materials (whether a single item or a boxful) are requested for reading room use, circulation, reproduction, or interlibrary loan, digitize them and make the digital versions available to everyone, by dropping them into the collection- or series-level descriptions. In this way, our initial selection of collections combines with their selection of items to float materials in likely demand to the surface.

C. For collections receiving minimal description or that are in the backlog, scan signposts and then devote more effort as use and interest warrant. The Greene/Meissner study[2] persuaded us to expose tips of the icebergs, by means of multilevel descriptions and minimal processing. This approach can be extended to digitization. Quick, representative scans can bring our collections out of the shadows and into the light. Both scholars and casual searchers can see the representations of some parts of the collection on the web. Some might choose to follow up to figure out what lies below the surface of the web.

We've already done the selection. Stop making decisions. If it's not valuable, why do we have it? If it's useful, researchers of all stripes may want it -- but only if they know it's there.

3. Do it ONCE (then iterate)

Don't get further behind. Do the scanning as an integral part of the initial accessioning and processing, rather than taking extra time to make another pass. Don't let newly acquired collections enter directly into the backlog.

We can make some compromises on image resolution and detailed metadata for the scans. High-level description with some representative scans may be enough (and is certainly better than keeping the collection off-limits, waiting for the rainy day that may not come). Make scanning and processing a collection into a single unified process, so you only need to touch it once.

However, once you've got all your collections represented on the web, you now have a basis upon which to determine where to apply further effort. Examining use is a great way to learn

about researchers' needs. Trueswell's 80/20 observation suggests that 20% of library materials will satisfy 80% of requests. Iterate once you've identified that 20%.

This iterative process will ensure that any extra effort is warranted. We can devote time to more meticulous metadata as appropriate. We'll go back to scan more or at higher quality when it's justified.   Lather, rinse, repeat.

4.   Programs not projects

We need to provide, on the Internet, what researchers want and use. To do this, we need to embed this mission-essential commitment to digitizing archival materials throughout our libraries, archives, and museums.

Early digitization projects have focused on cherry-picking for topical projects. Now we find our vast collections represented by a relatively small number of gorgeous images, lovingly selected, described, and presented in deep web portals. These hand-crafted digital presentations don't begin to reflect the breadth of materials in our institutions. Digitizing our rare and unique materials must be integrated into the operating budget. Like processing. Like conservation. Like creating catalog records or finding aids.

Greater access is a significant part of our professional responsibility. Digitization is an important part of access. All other important processes have budget, staffing and infrastructure, why should digitization be different?

To do a better job of providing access to our collections we must integrate digitization into all workflows and user services. To help establish priorities, those describing the collections should engage in public service discussions, put in some time on the reference desk, and analyze requests received via e-reference. We should embrace collaboration at all stages in the research process. We can work with our researchers and curators and archivists and librarians and registrars and IT at every step to plan, implement, and operationalize the digitization of special materials.

Increasing access to special collections needs to be programmatically embedded across the enterprise. Continuing to give these activities "special project" status implies that providing access is not mission-essential.

5.   Describing special collections: Take a page from archivists

Stop obsessing about items. Everything that is digitized does not need to be painstakingly described. Archival control distinguishes organic collections from description of distinct books and museum objects. Let's embrace that collection management strategy as well as the standards and practices of managing collections and hierarchies, not necessarily items.

This practice can be effective for archive, library, and museum collections, whether they consist of manuscripts, photographs, or specimens.   Start at the top, at the collection level, then think how to group materials, stressing the *relationships* between the materials.

While serious researchers value the description and organization that we bring to collections, after they've discovered a useful resource, they're willing to contact us should they need more detail.

We should consider opening our finding aids and other descriptions for user contributed amendments and commentary. They may be willing to make corrections, expand the descriptions, even provide transcriptions of handwritten materials.   We could consider asking researchers using unprocessed or minimally processed collections to pitch in; they may know more about the content than we do.

We should also think about the way we present our descriptions. We keep treating our bibliographic records like catalog cards and our finding aids like type-written documents. Our practices may be meaningful for us, but they are not intuitive to users. The content needs to be presented in new ways to be more helpful to our online users.

6.   Quality vs. quantity—Quantity wins!

Well-intentioned efforts to help the community have contributed to the dire situation where special collections risk marginalization.   RLG and others have stressed capture and description at the highest quality-level possible. Funders have been drawn to support boutique curation of compelling collections. None of these approaches allow primary sources to have the significant exposure to users that they so richly merit.

Vast quantities of digitized primary materials will trump a few superbly crafted special collections. Minimal description will not restrict use as much as limiting access to those who can show up in person. We must stop our slavish devotion to detail; the perfect has become the enemy of the possible.

Instead of focusing on measuring the output of catalogers, we should put more effort into measuring the impact on users. Once we know which materials are in demand, we can devote our efforts to more exhaustive scanning, higher quality scans, enhanced metadata, and text conversion -- on the content that will most benefits our users.

7.   Discovery happens elsewhere

Our intricate attempts to describe and present a few choice collections have resulted in expensive, but little-used websites. And the rest of our collections remain largely invisible.

We need to stop thinking of our lovingly crafted sites, designed specifically for a particular collection, as the only way people will discover our content. While researchers value the description and organization that we bring to collections, they don't want to have to consult dozens of specialized sites to find what they need.

Usage will increase when we take a small, but critical, extra step to expose the content to search engines and aggregators, who then make it accessible in places where users are more likely to find it. And given that users who discover our materials in this way are likely to drop into the middle of a collection, we need to make it easy to link to contextual information.

We will not be the only distributors of our content. Our emphasis should go to making our content harvestable, collectible, and indexable by others who are more successful at reaching broad audiences. We should redirect some of our efforts from building stand-alone sites to finding out where our users spend their time and ensure that our materials are discoverable in those locations.

8.   Brother can you spare a dime?

All digitization and description is expensive and there is so much to be done. As a community, we need to find ways to get funding -- without selling out.

After the lower hanging fruit, the books, are done, private companies will probably be increasingly interested in forming partnerships to digitize special collections. Perhaps they could take a particular format (like photos, maps, manuscript pages) and develop scanning stations and industrial strength workflows that suit the format and function at scale. They could help us crawl out of the (book) cradle and start playing in the (special collections) sandbox.

However, we need to ensure that in any public/private arrangement, our needs are given as much weight as our partner's desires. NARA has set a good example by identifying internal principles to serve as guidelines in negotiation agreements.[3]

We should also ensure that the partnerships we enter into allow us to make the resulting content openly accessible so that we don't preclude important uses and functionality that will be desired by future scholars.

Getting grant funding in the past seemed to require identifying a "sexy" collection and maximizing its specialness in the ways we described, digitized, and presented it. A lot of money went toward creating barely visited web sites. And a lot of institutions created preservation-quality images that they, in fact, had no way to sustain in the long run.

Grant-giving agencies should be encouraged to support projects that put permanent processes in place for ongoing operations. They might follow NHPRC's lead in sponsoring initiatives that help to establish scalable processes that can be used again and again.[4]

But most importantly we all need to stop thinking in terms of individual projects and instead think in terms of ongoing mission-essential operations.

All these measures will help us to begin to keep pace with mass digitization of books. As we increasingly share a collective collection of books, it will be our special collections that distinguish cultural repositories. We can hide those treasures in our backlogs or behind custom portals -- or we can push them out into the light of day.

Notes

[1] Hafner, Katie. 2007. "History, Digitized (and Abridged)." Your Money, *The New York Times,* Business section, March 10. Available online at: http://www.nytimes.com/2007/03/10/business/yourmoney/11archive.html [accessed 28 September 2007].

[2] Greene, Mark A., and Dennis Meissner. 2005. "More Product, Less Process: Pragmatically Revamping Traditional Processing Approaches to Deal with Late 20th-Century Collections. *American Archivist,* 63,2 (Fall/Winter). Pre-print available online at: http://ahc.uwyo.edu/documents/faculty/greene/papers/Greene-Meissner.pdf. [Accessed 11 October 2007.]

[3] National Archives and Records Administration. 2007. *Plan for Digitizing Archival Materials for Public Access, 2007-2016."* Draft version of 10 September, available for public comment online through 9 November 2007 at: http://www.archives.gov/comment/nara-digitizing-plan.pdf. [Accessed 11 October 2007.] See Appendix A.

[4] National Historical Publications and Records Commission. 2007. "Strategies and Tools for Archives and Historical Publishing Projects." Grant announcement. Available online at: http://www.archives.gov/nhprc/announcement/strategies.html. [Accessed 11 October 2007.]

Appendix

Digitization Matters: Breaking through the barriers

Scaling up digitization of special collections

On August 29, 2007 RLG held a forum, co-sponsored by SAA, and hosted by the Newberry Library. The agenda was to discuss how to advance digitization of primary sources, in light of the recent efforts toward mass digitization of books. Initially it was hoped that 60 people would register; registration was closed after two hundred registrants. It was clearly a timely topic.

The audience included directors, administrators, and curators of special collections in libraries, archives, and museums. We asked attendees to join us as we tempered our historical emphasis on quality with a recognition of the need for quantity—and suggested that they come with an open mind. We've asked a variety of speakers to go out on a limb, to think wild thoughts, to speak not from their institution's point of view, but to act instead as a catalyst for discussion. After each set of provocative ideas, the audience engaged in active discussions in the hopes of surfacing several suggestions that were implementable and identifying others that might benefit from further investigation. Surprisingly, those ideas that were meant to provoke met with very little resistance, but were instead welcomed as new ways to approach the road ahead.

The speakers were:

- Susan M. Allen, Getty Research Institute
- Susan Chun, Consultant
- James Eason, Bancroft Library, U.C. Berkeley
- Sharon E. Farb, UCLA Libraries
- James J. Hastings, National Archives of the United States
- Michael Jenkins, Metropolitan Museum of Art
- Bill Landis, Yale University Library
- Barbara Taranto, New York Public Library
- Sam Quigley, Art Institute of Chicago
- Günter Waibel and Ricky Erway, OCLC Programs and Research (Moderators)

The speakers' ideas and the audience discussion provided the fodder for the essay, "Suggesting a Shift," by Ricky Erway and Jennifer Schaffner. We're grateful to the speakers for starting this important discussion and to the audience for enthusiastically taking it forward. If you would like to listen to the forum, digital audio recordings are available in mp3 format at http://www.oclc.org/programs/events/2007-08-29.htm